Academia Journal of Scientific Research 8(2): 090-094, February 2020
DOI: 10.15413/ajsr.2019.0176
ISSN 2315-7712

Academia Publishing

Research Paper

Characteristic assessment of advancement in the role of information professionals and libraries in data science era

Accepted 11th November 2019

Naimat Ullah Shah*, Salman Bin Naeem and Rubina Bhatti

Department of Library and Information Sciences, Islamia University, Bahawalpur, Pakistan.

*Corresponding author's email: naimat784@gmail.com. Tel: +923327775222

ABSTRACT

Data sciences usually involves data management, its utilization, distribution as well as its re utilization. All these components need to be focused while targeting data science. Thus data put a great burden on research institutes because they are the authority that is responsible for whole course of procedure. It is of prime importance for data science information professionals serving in data centric age to know about LIS principles, theories and other related skills that are mandatory for management and support of data science. This study sums up the reviews of researchers regarding data science era. Moreover, this study includes characteristic assessment of data science environment with respect to recent advancements in data science and advancement in duties of librarians, presentation of classified data, function of data science libraries as well as librarians with respect to data users. It is supposed to be interesting era to work in a library as its role is expanding with certain new challenges. There is need for the current era to educate librarians, library science researchers and students regarding understanding, utility and management of data to meet the requirements of data science librarians.

Key words: Data science, data management, data librarians, data managers, data scientists, data centers.

INTRODUCTION

Data science is involved in all stages in university libraries and data-centric studies are perceived as a usual attachment of digital data-management and electronic resource tasks but the main thing is the level of awareness and skills on how to know if domain considerate is needed (, 2008). This 'data centric era' requires universities to establish strategies, policies, setups, and data centric services for data-management, through assisting the scholars in establishing, gathering, using, assessing, managing and sharing the digital data-sets (Pinfield et al., 2014).

Certain nascent research corners are unveiled every decade. Data analysis and computer sciences are going to make us witness some quickly adopted revolutionary advancements. With the appearance of nascent areas of research, certain novel terminologies also come into play such as data mining, deep learning etc. In a broader sense, these are actually certain generalization of existing fields. With the significance of such generalizations, not only the appearance of new terms happens but it also serves as augmentation impulse for the quick progress of the field. Data science, being one of the newly established terms, is looking for more targeted and purified definition (Dzemyda, 2018). The notion of data is the basis on which the entire building of data science relies. Data scientists have concluded that data are not only goal oriented but also neutral in itself. Information studies have defied this perception by regarding data to be bias from inside. Scientific scholars have also concluded the same. The word data is actually described as the statistical observations and findings or collection and recording of proofs or a succession of detached findings and observations (Zins,

2007).

Objectives

- To know about the characteristic assessment of data science environment with respect to recent advancements in data science through review of related literature.
- To know about the role of information professionals in data science era through review of related literature.
- To know about the advancement of digital services in libraries in data science era through review of related literature.

METHODOLOGY

This study summarizes the reviews of researchers regarding data science era. Moreover, this study provides theoretical lens which includes characteristic assessment of data science environment with respect to recent advancements in data science and advancement in duties of librarians, presentation of classified data, function of data science libraries as well as librarians with respect to data users.

LITERATURE REVIEW

In a broader sense, the data may be defined as the assortment of digits or sequence of text or a thread of alphanumeric codes which will not necessarily possesses a firm meaning. Data itself rarely holds a noteworthy value (Frederick, 2016). Data are an element or bit of information which have the observed volume to be gathered and kept in a way to be recognized, used again and again. "Data are famously divided into two types: structured-data and unstructured-data (Ramkumar, 2018).

Data are of prime importance in research activities because of its towering worth in information. Actually, data regards information to possess a level of significance that may be considered as critical for making a decision on a certain matter. Data are supposed to appear in a variety of ways including surveys, statistical declarations, research outcomes, graphical presentations, figures or illustrations, interviews, journals statements etc. (Mikalef et al., 2018). Data are considered as a vital factor for science and technology. They play a key role in generating hypothesis, assessment and making desired amendments in a hypothesis, putting forward theories and suggesting models. Data have basically made it possible for technology to be practiced and show applications in day to day life (Oyelude, 2017).

Novel ways of data implication

Presently, the research studies are more data centric and scholars are facing novel type of barriers in accessing, disseminating and management of digital data, therefore, libraries are beginning to offer data services, which includes training and instructing, management of data, planning and data guidance, data-curation and stewardship, and data imagining (Federer, 2018). Science is not only the owner and creator but also the managing authority of the biggest share of universal data. General opinion of public is that data are brought into existence by scientist community for scientific purposes. However, if we talk about data with respect to digital stuff, it may be utilized for multiple intellectual goals (Clement et al., 2017). Similarly, the Data may also be employed for production of nascent ideologies, novel protocols as well as invention and deployment of scientific information in novel ways for the significantly improved customary computational, experiment based, physical observations based as well as theoretical tracks for the discovery of new scientific stuff and their utilities for public. The procedure can initiate generation of idea followed by scheduling of track for data collection as well as its storage for future utilities. In academia, such initiation involves advance proposal as the first step. Individuals from scientific backgrounds are usually not well trained regarding the preservation techniques of data for longer periods (Cao, 2018).

Historical development of data science

Data Science has been specifically acknowledged in the present era as an area of study that deals with huge data stores. As far as data ae concerned, a huge variety of perceptions are suggested by data scientists and experts of its allied areas including statistics, library science and computer science long ago. Back in the late 5th decade of the 20th century, a nascent term Bit was introduced by Turkish scientist and was phrased by Claude Shannon in one of his papers (Shannon, 1948). After being presented in a study "A Mathematical Theory of Communications" by Claude Shannon, another researcher John W. Tukey mentioned it in his paper "The Future of Data Analysis" (Tukey, 1962).

In the year 1977, Tukey put a huge emphasize on the utilization of data in various phases of research including hypothesis, its assessment and final analysis (Tukey, 1977). The *International Federation of Classification Societies (IFCS)* biennial conference held at Kobe, Japan became the pioneer who uses Data Science term as conference title (Davenport and Patil, 2012). It is supposed that the initial introduction of data science as a term in the official documents and manuscripts was in 1974 in preamble to Naurs book "Concise Survey of Computer Methods" (Belzer, 1976). In the intro of that book, the definition of data science is mentioned that science involves dealing with the data just after their establishment as well as relating the data along with their findings and presentations to other sub fields of data science. Word data science was coined by statisticians academically to point out the subject as "big

data, data analysis, and broader trends", which highlights statistical foundations and the new mathematical techniques to make sure of the richness of data (Gilmore, 2016).

Data science era

The greatest challenge for science in twenty-first century is how to report to the new data science era. In experimental and theoretical research, data science is recognized as a emerging model and also in computer and information science recreations of natural phenomena which is require new ways, techniques and tools of working (Peyne and Chan, 2017). Data Science, being also known as e-science, is regarded as a combination of scientific procedures and technologies that assist in data alliance and collaboration for making it possible to assess as well as explore the data along with supporting communication and propagation between the scholars (Antell et al., 2014). Scientific data is mainly composed of each and every corner of natural, artificial and socials systems, all among which act as source for obtaining scientific data. The data obtained from these components or sources are passed via a chain of scientific protocols to get the desired goals. The scientific data are a source of immense opportunities for the data sciences, introducing it to a variety of new prospective (Federer, 2016). The data have to be transformed into a compatible and useful format before exposing them to be processed by a computer. This is regarded as the equation's input portion. It is preferred to be obtained in a useful context. This is regarded as the equation's input portion. The output is supposed to be associated with input, as suggested by adage, "Garbage in, Garbage out" (Balachandran and Kamalanathan, 2018).

It is obvious that still the importance of data science and its analytical approaches in enabling of data driven theory, economical progress and professional improvement is going to be accepted with much more potential than ever. This is composed of not only the basic fields of computer, information technology and statistics but also certain other disciplines such as business, social sciences and health sciences (Cao, 2018). The establishment of a systemic educational model is mandatory to train our future's data professionals (including engineers, professionals, scientists as well as executives) which may allow them capable of having intellectual attachment with data, gain expertise regarding management of data, have a hand on computation of data, gain abilities to mine data, successfully propagate data, know about the delivery of data and know how and when to take action on data (Waller and Fawcett, 2013). It is an approach to do research as studies of "big data, predictive analytics and data science" which is the requirement of researchers with the field of information and data management (Carillo, 2017)

Library and information science in data science era

The data-science can only be casted as a discipline if a specified foundation is established for it. Moreover, advancements of technological fields, specification of subject areas, development of research map and goal and establishment of practical tools are also considered as vital factors for such purposes.

The progress of information science research and education is going on with such a pace in LIS schools that can be regarded as the leading reign of development of these fields. Keeping in view the challenges faced by and opportunities provided by data science, it is obvious that information science will pay incredible gifts to the field of data science and collaboration of these two will certainly benefit the society (Wang, 2018). Data science, along with a huge extent of data and its applications, has become known as a nascent field in present decade. A number of i-schools have been established with fewer schools of library and information sciences (Zuo et al., 2017). Data science involves a variety of "Data Issues" that resides in cybernetics, information theory and system sciences which are supposed to be customary basis of information system and convert them into a logical and theoretical idea. So data science is going to replace the customary theories into a nascent theory source of IS presenting novel speculative references for it (Ge et al., 2018). "Data chain and information chains are intrinsically associated with each other. Basis of both of them relies on extracting value as well as insight from data" (Cai and Zhu, 2015). Data science perfectly incorporated with information science. We may conclude that data can be proved as a research object for information sciences. Moreover, both may have tremendous alike attributes. However, massive data never always represents a huge potential for LIS School. In case we manage to successfully prove the association and common aspects in data and information studies, it is necessary to propose data studies which may become component of information science and may lead to a very effective mode of education in LIS schools (Wang, 2018).

Role of information professionals in data science era

The role and responsibilities of Information Professionals have greatly changed. The involvement of data science in data management libraries has become an integral portion of research protocol from the very start of life cycle (Walek, 2017). The nascent job title of data librarian has come into existence in such a data science scenario. The initiation of data librarian may be linked to social sciences, bioinformatics and data management research activities (Koltay, 2017). Currently, the progress of data library professionals particularly those related to academic and research, is seen to be very fast. Data librarian is actually produced by intellectual broadening of the term of

academic librarian, keeping in view the new requirements, research policy needs and nascent data management performances (Federer, 2016).

According to Lyon and Mattern (2017) stated that "*There are six roles of data science managers they are: data archivist, data curator, data librarian, data analyst, data engineer, and data journalist*." The mentioned six characters are enclosed as data-science characters, some of the other roles are derived by the commercial site, which define analysts of data kind job oriented as data-scientists.

Data Science works as a key element in information centers, so building librarians with data-science techniques enhance their skills to keeping up these data-centered studies (Burton et al., 2018). Article on data has been known since earliest time, which may experience tests, peered review, and circulation, use again and again, and is mainly under consideration. The data reuse helps the society to reuse an idea in a journal article. The libraries and its staff have been found to pay duties of collection, preservation and dissemination of intellectual results of the society (Carlson and Johnston, 2015). Previous studies have shown absence of an agreed guideline regarding duties and requirements with respect to data science qualification, skills and capabilities that are supposed as prerequisites for the position of "Data Librarian" (Yoon and Schultz, 2017). Librarians should also be skillful in handling and assessing technologies for future data-science era. For usual provision of services through technology, literature suggests that library information science may require specific care in outgoing and gaining undesirable approaches towards technology (Lyon and Mattern, 2017).

Functions of libraries in data science era

With the progressive involvement of digital data dealing, the libraries have either made the boarders of professional duties of their existing staff wider by making them endow data services to their concerned groups or have managed to appoint full time and committed additional professional staff that has expertise in rendering data services (Lyon, 2012).

Libraries are the sites that are expected to possess the skill sets, durability and the vital skeleton for accomplishment of tasks for a variety of data. In case of negligence from libraries in engagement in such tasks, the public will alternatively originate certain new institutional setup to deal with digital data. This situation will mainly direct us towards the textual output's and related data's dissociation that will unfortunately result in deprivation of society to benefits from its own knowledgeable assets (Heidorn, 2011). Libraries usually possess an attraction for data. Almost the men's entire actions make libraries to need data for accomplishment of their technical tasks. In absence of data, the needy person will not be able to meet the desired information. Circulation of books and payments of staff may also be disturbed (Frederick, 2016). Current Emerging technologies in stats and software engineering tethered with richness of information have provides a novel expert bio network known as data-science. Data-science technique can changed trade, health science, and management and it can also change libraries like other areas.*"There is the need of Investments to enlarge the recent flow of support in the area of data science"* world to more and more imbued with information; information professional was a critical job in growth and future of the data-science (Burton et al., 2018). Data science's fundamentals specifically theory basis and procedures are used for analyzing and gaining knowledge towards certain corners including numerical computation, algebra, geometry, graph, probability, and theory of information science (Van Der Aalst, 2016).

CONCLUSION

The data scientists are not likely to come from conventional library backgrounds, predominantly career researchers who are gone via a period of work as a data scientist as a mandatory part of long term research career track. Who will possibly provide such advice? It is possibly the data librarian serving in university, who may also accomplish a task of administration of local data sets for tiny projects, the need that may be held in case the huge gaps in national level stipulation are stopped in the future.

There is the need of the current era to educate librarians, library science researchers and students regarding understanding, utility and management of data. Science is not only the owner and creator but also the managing authority of the biggest share of universal data.

There is a need of current scenario that information professionals inquire what data science roles can fill out through current assets by librarians. Libraries are in a good position to contribute to providing basic infrastructure needed for support of data science, particularly admittance to data, utilization, reuse, storage and preservation of data. It is supposed to be interesting in data science era to work in a library as its role is expanding with certain new challenges.

REFERENCES

Antell K, Foote JB, Turner J, Shults B (2014). Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. Coll. Res. Libr. 75(4): 557-574.

Balachandran K, Kamalanathan S (2018). Data Management in Clinical Research. Thesis Writing for Master's and Ph. D. Program. pp 83-92.

Belzer J (1976). Concise survey of computer methods. Peter Naur. New York: Petrocelli Books, 397 p.(1975). J. Am. Soc. Inf. Sci. 27(2): 125-126.

Burton M, Lyon L, Erdmann C, Tijerina B (2018). Shifting to Data Savvy: The Future of Data Science In Libraries.

Cai L, Zhu Y (2015). The challenges of data quality and data quality assessment in the big data era. Data Sci. J. 14: 2. DOI: http://doi.org/10.5334/dsj-2015-002.

Cao L (2018). Data Science Thinking: The Next Scientific, Technological and Economic Revolution: Springer.

Carillo KDA (2017). Let's stop trying to be "sexy"–preparing managers for the (big) data-driven business era. Bus. Process Manag. J. 23(3): 598-622.
Carlson J, Johnston LR (2015). Data information literacy: Librarians, data, and the education of a new generation of researchers (Vol. 2): Purdue University Press.
Clement R, Blau A, Abbaspour P, Gandour-Rood E (2017). Team-based data management instruction at small liberal arts colleges. IFLA J. 43(1): 105-118.
Davenport TH, Patil D (2012). Data scientist. Harvard Bus Rev. 90(10): 70-76.
Dzemyda G (2018). Data Science and Advanced Digital Technologies. Paper presented at the International Baltic Conference on Databases and Information Systems.
Frederick DE (2016). Data, Open Science and libraries–The Data Deluge Column. Library Hi Tech News. 33(8): 11-16.
Ge M, Bangui H & Buhnova B (2018). Big data for internet of things: A survey. Future Gener. Comput. Syst. 87: 601-614.
Gilmore RO (2016). From big data to deep insight in developmental science. Wiley Interdisciplinary Reviews: Cogn. Sci. 7(2): 112-126.
Heidorn PB (2011). The emerging role of libraries in data curation and e-science. J. Libr. Admin. 51(7-8): 662-672.
Henty M (2008). Dreaming of data: the library's role in supporting e-research and data management.
Koltay T (2017). Data literacy for researchers and data librarians. J. Librariansh. Inf. Sci. 49(1): 3-14.
Lyon L (2012). The informatics transform: Re-engineering libraries for the data decade. Int. J. Digit. Curation, 7(1): 126-138.
Lyon L, Mattern E (2017). Education for Real-World Data Science Roles (Part 2): A Translational Approach to Curriculum Development. Int. J. Digit. Curation, 11(2): 13-26.
Mikalef P, Pappas IO, Krogstie J, Giannakos M (2018). Big data analytics capabilities: a systematic literature review and research agenda. Inf. Syst. e-Bus. Manag. 16(3): 547-578.
Oyelude AA (2017). Data, data everywhere, but little to use! Library Hi Tech News. 34(8): 24-25.
Peyne B, Chan A (2017). Data-driven decision making in Marketing: A theoretical approach.
Pinfield S, Cox AM, Smith J (2014). Research data management and libraries: relationships, activities, drivers and influences. PLoS One, 9(12): e114734.
Ramkumar P (2018). Initiatives of Research Data Literacy for Data Management in Libraries. J. Adv. Libr. Inf. Sci. 7(1): 128-131.
Shannon CE (1948). A mathematical theory of communication. Bell Syst. Tech. J. 27(3): 379-423.
Tukey JW (1962). The future of data analysis. Ann. Math Stat. 33(1): 1-67.
Tukey JW (1977). Exploratory Data Analysis: Limited Preliminary Ed: Addison-Wesley Publishing Company.
Van Der Aalst W (2016). Data science in action Process Mining Springer. pp. 3-23.
Walek A (2017). Is data management a new "digitisation"? A change of the role of librarians in the context of changing academic libraries' tasks.
Waller MA, Fawcett SE (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J. Bus. Logist. 34(2): 77-84.
Wang L (2018). Twinning data science with information science in schools of library and information science. J. Doc. 74(6): 1243-1257.
Yoon A, Schultz T (2017). Research data management services in academic libraries in the US: a content analysis of libraries' websites.
Zins C (2007). Conceptual approaches for defining data, information, and knowledge. J. Am. Soc Inf. Sci. Technol. 58(4): 479-493.
Zuo Z, Zhao K, Eichmann D (2017). The state and evolution of US iSchools: From talent acquisitions to research outcome. J. Assoc. Inf. Sci. Technol. 68(5): 1266-1277.

Cite this article as:

Shah NU, Naeem SB, Bhatti R (2020). Characteristic assessment of advancement in the role of information professionals and libraries in data science era. Acad. J. Sci. Res. 8(2): 090-094.

Submit your manuscript at

http://www.academiapublishing.org/ajsr